Remote Analysis of Climate Model Output

Velimir Mlaker[1], Dean N. Williams[1], Roland Schweitzer[2], and Steve Hankin[2]
[1]Lawrence Livermore National Laboratory (LLNL) and [2]Pacific Marine Environmental Laboratory (PMEL)

Summary

Climate models are generating ever-increasing amounts of output data. The growing size and complexity of these large datasets makes it increasingly cumbersome for scientists to stage in-house analysis of the simulation results. For these and other reasons, server-side analysis of climate model results is quickly becoming a necessity. Universal web-based access to data and scientific tools is one of the goals of the Earth System Grid Center for Enabling Technologies (ESG-CET). As part of this "*science gateway*", the Live Access Server (LAS) will provide scientists with a seamless interface to the Climate Data Analysis Tools (CDAT), a powerful suite of computational tools designed to work on the demanding volume of next generation climate model output.

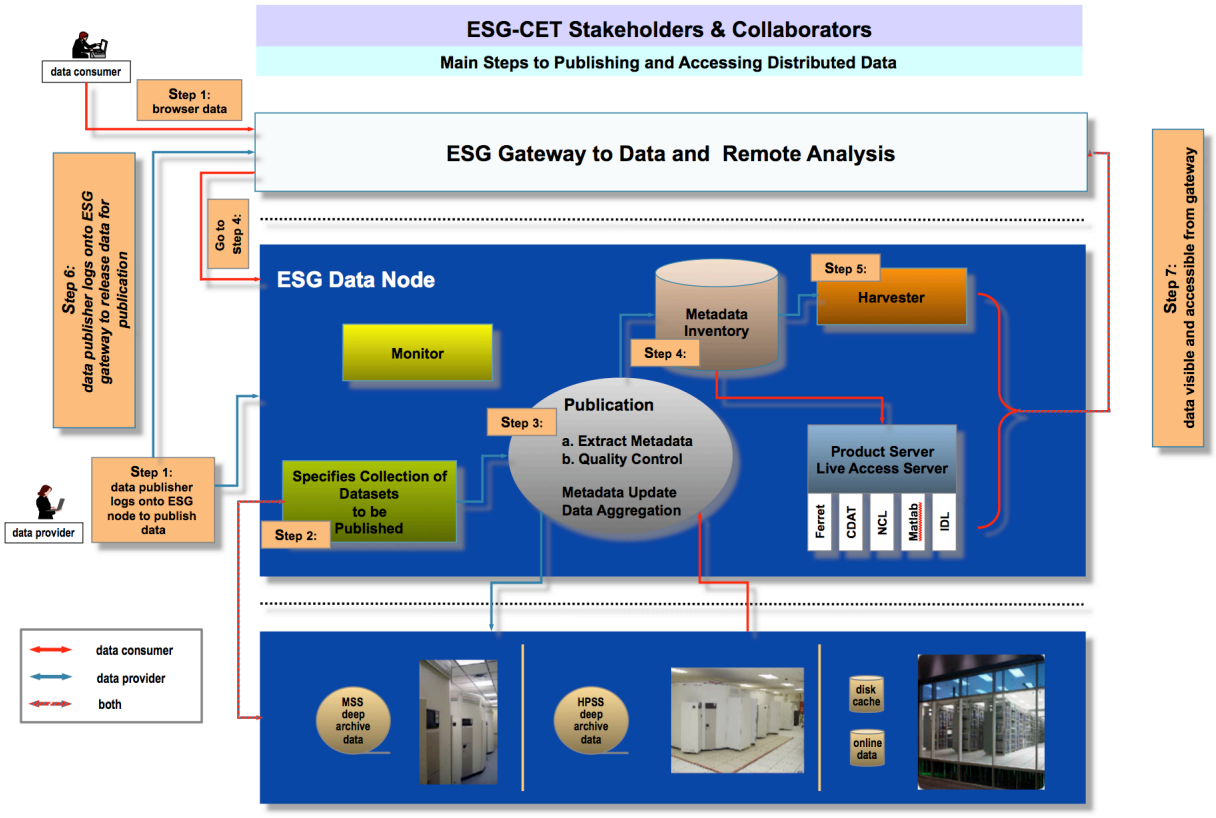

Figure 1. The "*Data Node*", together with the Gateway forms the two main components of the ESG architecture. Data Nodes are hosts where data is actually stored or archived. As this figure illustrates, the functionality of the Data Node is to publish data (making it visible to an ESG Gateway), and provide data analysis and delivery services to ESG end users. With CDAT and other analysis capabilities (such as Ferret, NCL, etc.) in LAS, climate scientists on the ESG grid can transparently access large datasets from disparate data centers and perform complex, configurable, ESG "*Data Node*" operations in a lightweight user interface environment.

Every new generation of global climate models generates far greater amount of data than the previous. For the Coupled Model Intercomparison Project, Phase 5 (CMIP5), total output is expected to be on the petabyte scale. The huge size of these datasets presents challenges to the community of climate scientists tasked with analyzing them. Bandwidth limitations of network transfer will render traditional client-side computational tools inadequate for accessing and analyzing this volume of data. It will become increasingly prohibitive to download frequently updated data and run analysis in-house. Instead, the paradigm will increasingly shift to performing analysis at the server where the data reside.

As modeling groups publish updates or corrections to their output, integrity of data becomes an important factor in scrutinizing provenance of analyses. Server-side analysis will become increasingly necessary to certify validity of data. Analyzing at the server will guarantee that operations are performed on the most up-to-date model output.

In addition to benefits of access efficiency and integrity inherent in server-side analysis, scientists will benefit from computational power of the server. The most advanced analysis software tools will run on the latest enterprise scale hardware. This will give users access to high-end computing platforms that may otherwise be out of their reach.

With ESG-CET, these goals are being met by implementing server-side analysis using the Live Access Server (LAS) and the Climate Data Analysis Tools (CDAT) working in concert on the ESG-CET Node (see Figure 1).

The LAS is a highly configurable web server designed to provide flexible access to geo-referenced scientific data. Developed at the National Oceanic and Atmospheric Administration (NOAA) Pacific Marine Environmental Laboratory (PMEL), it offers a web browser front-end utilizing Web 2.0 technology that is designed to allow efficient and scalable user interface for exploration, subsetting, comparison, and visualization of datasets in a platform-independent way.

By combining intuitive end-user interface with powerful analysis features, remote analysis on the ESG-CET "*Data Node*" will be available instantaneously to a worldwide audience. "*Data Nodes"* serves data and data products securely and reliably to the end users. "*Data Nodes"* are application servers co-located with the data, which can be flexibly configured depending on the specific community needs and hardware and administrative resources (see Figure 1).

The actual analysis functionality will be handled by web services implemented in today's commonly used analysis tools such as: Ferret, CDAT, and NCL. For this research, CDAT is of particular interest. It is a powerful and highly extensible climate data analysis suite. CDAT is developed at PCMDI at LLNL. It is open-source and has an extensive user community that continuously contributes useful application code, making CDAT an ideal choice for analysis of next-generation climate model output.

Integrating the feature-rich functionality of CDAT into the scalable and extensible server architecture of LAS will offer users highly available one-stop access to climate data analysis. The ESG-CET analysis framework will combine benefits of data integrity, network efficiency, and computational power, thus alleviating the scientist of the burden of staging complex workflow at the client-end, which is time consuming and error prone. The ease-of-use of this framework, together with transparency of the underlying processes and technology, will invite non-experts into the domain of climate model output analysis.

For further information on this subject contact:

- Name: Dean N. Williams
  Organization: PCMDI, LLNL
  Email: williams13@llnl.gov
- Name: Roland Schweitzer
  Organization: PMEL, NOAA
  Email: Roland.Schweitzer@noaa.gov